Towards a generic compilation approach for quantum circuits through resynthesis

Arianne Meijer - van de Griend

***Abstract*—In this paper, we propose a generic quantum circuit resynthesis approach for compilation. We use an intermediate representation consisting of Paulistrings over $\{Z, I\}$ and $\{X, I\}$ called a "mixed ZX-phase polynomial". From this universal representation, we generate a completely new circuit such that all multi-qubit gates (CNOTs) are satisfying a given quantum architecture. Moreover, we attempt to minimize the amount of generated gates.**

We show that for large circuits, containing ≥ 100 Paulistrings, our algorithm generates significantly less CNOTs than previous methods, as well as both the TKET compiler and Qiskit transpiler.

Note for QPL reviewers: when implementing PermRowCol, I realized that my implementation of the paritysynth [24] did not properly correspond to the pseudocode and I have changed it. As a result, the algorithm is much slower, taking more than 2 minutes on average to compile a 20 qubit circuit with 100 gadgets for the target devices. This means that running the script that compares the different methods takes approximately 29 hours. Because this is on the brink of what is acceptable runtime, I have implemented my previous algorithm [16] to compare against. It is much faster but less CNOT-efficient than paritysynth (see also [24]). I am now rerunning all experiments and expect the results on Wednesday morning. I have updated the tables with my latest results and also included the runtimes. The text does not yet fit these new findings because they are still not final.

I. Introduction

Small scale quantum computing has recently become accessible to the public through cloud APIs like AWS Braket [2], IBM Qiskit [1], among others. With these new computing platforms comes the need for new compilation methods for quantum programs. Although some classical compilation methods can still be used for quantum computers, some compilation problems require a new solution due to the quantum nature of these devices. For example, most devices do not allow operation between arbitrary qubits. This is generally true of classical computers as well, but there we can solve the problem by copying bits. However, copying qubits is not possible due to the no-cloning theorem [27]. Similarly, optimizing quantum programs is a complex task. We somehow need to understand what the program is doing without accidentally executing it. Because simulating arbitrary quantum programs is infeasible on classical computers. In this paper, we propose a new compilation method for quantum circuits that optimizes the quantum circuit, and in doing so, it generates only quantum operations as allowed by a target quantum computer.

We essentially do this by throwing away the original circuit and generating a completely new circuit. This means that we do not need to cut the circuit into small pieces (as in e.g. [8, 17, 11]) and we do not generate any SWAP gates explicitly (as in [12, 25, 18]). Hence, this is a full-stack re-synthesis approach. Moreover, this is the first resynthesis algorithm that compares itself to common compilers like Qiskit [1] and TKET [23].

Since this approach transforms quantum circuits into quantum circuit, it is technically a transpilation algorithm rather than a compilation algorithm. We will use these terms interchangeably.

Note that the global nature of our proposed approach makes it probably unsuitable as an intermediate step in an existing compilation, unlike for example peep-hole-optimization.

Our algorithm is a combination of three previous algorithms. We will describe these algorithms in more detail in the next section, but for the familiar reader, we give a short overview here. We start with representing a given circuit in an intermediate representation (IR). We use the mixed ZX-phase polynomial IR that was used in [9] which is essentially a sequence of Paulistrings either over the alphabet $\{Z, I\}$ or $\{X, I\}$. This is a universal representation. We synthesize the circuit using the current state-of-the-art phase polynomial synthesis algorithm [24] and optimize the generated CNOTs with PermRowCol [17]. Then, we used simulated annealing to simplify the IR by adding a few CNOTs to the circuit. This last step essentially removes common CNOTs from the original Paulistrings.

Note that Qiskit now uses SABRE

Unfortunately, due to the heuristic nature of our algorithm, we cannot make claims about the minimality of newly generated circuits. However, by comparing our algorithm to Qiskit, TKET, and it's subcomponents, we know that the generated circuits have significantly less CNOTs than previous methods.

Our approach is very powerful for circuits generated from multi-qubit Pauli-interactions such as those generated by transforming a Hamiltonian to a quantum circuit.

In the remainder of this paper we will describe quantum circuit re-synthesis problem in II, then we describe our algorithm in III. We compare the performance of the proposed method in IV. We give an in-depth discussion of the assumptions we made in the design of this algorithm and how one can or cannot deal with them (V). Lastly, we summarize and point to future research directions in VI.

II. Problem description

The problem that is addressed in this paper is commonly known as the "qubit routing problem" [5]. It is the problem that many quantum computers only support multi-qubit interactions between specific pairs of qubits. This means that not every arbitrary quantum program (quantum circuit) can be immediately executed on an arbitrary quantum computer. The quantum circuit will need to be changed first to accommodate these connectivity constraints (transpilation). Due to the "no-cloning theorem" [27], we are unable to reuse the existing classical methods because they require the copying of quantum states.

As the name suggests, the qubit routing problem is historically solved by routing the qubits: moving the qubits to different registers by applying SWAP operations until the connectivity constraints are satisfied [5]. However, each additional SWAP gate is generally implemented using 3 CNOT gates, which can generate significant overhead for the execution time of the quantum circuit. Since current state-of-the-art quantum computers are noisy, with low decoherence times, and sometimes sparsely connected qubits, the routing overhead can make a circuit impossible to execute.

Thanks to a recent paradigm shift, there is now a class of transpilation methods that do not rely on SWAP gates [20, 11, 8, 16, 17]. Instead, these algorithms rely on finding a scalable representation of the quantum circuit from which a completely new, but equivalent, circuit can be synthesized. During the synthesis process, only multiqubit interactions that are natively allowed are generated, hence the name "architecture-aware resynthesis" [16]. Most of these methods rely on the concept of "Steiner trees" to find how the existing connections can be optimally used. As such, this class of transpilation methods is sometimes called "Steiner-tree-based methods" [17]. An added benefit of resynthesizing the quantum circuit is that it will also be immediately optimized.

A. Mixed ZX-phase polynomial intermediate representation

Our method relies on the intermediate representation named the "mixed ZX-phase polynomial" [9]. We describe it here in detail and how it can be used for resynthesis. Unlike most existing literature, we will not rely on ZX-calculus [26] for this. It is true that mixed ZX-phase polynomials have a very intuitive representation in ZX-calculus, but its use is not yet mainstream enough to assume the reader's familiarity.

Given an arbitrary quantum circuit, we can rewrite that circuit into a sequence of multi-qubit Pauli operations $e^{i\alpha \otimes XYZI}$ called "Paulistrings". These operations are usually generated when transforming a Hamiltonian into a quantum circuit, and we assume the reader's familiarity with these operations. Note that a single-qubit gate can also be described as a Paulistring where every letter in the string is I except for one, which corresponds to the axis of rotation for that single-qubit gate.

A circuit consisting only of Paulistrings can then be rewritten into a sequence of Paulistrings either over the alphabet $\{Z, I\}$ (Z phase gadget) or $\{X, I\}$ (X phase gadget). Since Z-phase gadgets mutually commute and X-phase gadgets also, we can group the sequence into sequences of mutually commuting phase gadgets. A sequence of Z-phase gadgets is called a "phase polynomial" [3, 4], hence the name "mixed ZX-phase polynomial".

The advantage of the use of Paulistrings is that they can be efficiently represented. It is fully characterized by the string itself and its angle of rotation α. For the Z- and X-phase gadgets, the string can be represented as a bitstring (given that you know it is a Z- or X-phase gadget). Thus the original quantum circuit can be represented as matrix of $(n+2)\times m$, where n is the number of qubits ($+1$ for the angle, $+1$ for whether it is a Z- or X-phase gadget), and m is the number of phase gadgets in the mixed ZX-phase polynomial. Additionally, this representation behaves nicely when CNOTs are acted upon it.

Consider a circuit in which our intermediate representation is a black box. If two consecutive CNOTs were applied at the start of the circuit, the circuit would remain semantically the same. If one were to then move one of the CNOTs through the black box, the intermediate representation changes as follows:

- For **Z-phase gadgets**, the **target** qubit is added to the **control** qubit (modulo 2) in the bitstring.
- For **X-phase gadgets**, the **control** qubit is added to the **target** qubit (modulo 2) in the bitstring.

And the CNOT that was "pushed through" is now behind the black box.

If, at any point during this process, there is a single qubit phase gadget in the *first* sequence of mutually commuting phase gadgets, that phase gadget can be removed from the intermediate representation, and added as a single qubit gate just before the black box.

See Figure 1 for a visual representation of this process. The multi-qubit gate represents the black box with the mixed ZX-phase polynomial representation inside it.

The key observation for this approach to resynthesis is that we need to pick CNOTs that are allowed by the architecture AND optimally reduce the complexity of the intermediate representation in the black box. We want to take advantage of the CNOT being pushed through when it changes the bitstring representations of each phase gadget such that it requires fewer CNOTs later.

In the next section, we will describe how we did this in more detail.

III. The algorithm

Once the circuit is in the mixed ZX-phase polynomial representation, we need an approximately optimal strategy to synthesize the circuit from it. In this paper, we have have combined three previously proposed algorithms [9, 17, 24] into a single procedure. The aim is that these three algorithms work together in such a way that the whole performs better than the separate parts.

First, we use the simulated annealing approach proposed by [9]. The original algorithm adds pairs of CNOTs to the front of the circuit and pushes one of them through the intermediate representation, as explained in the previous

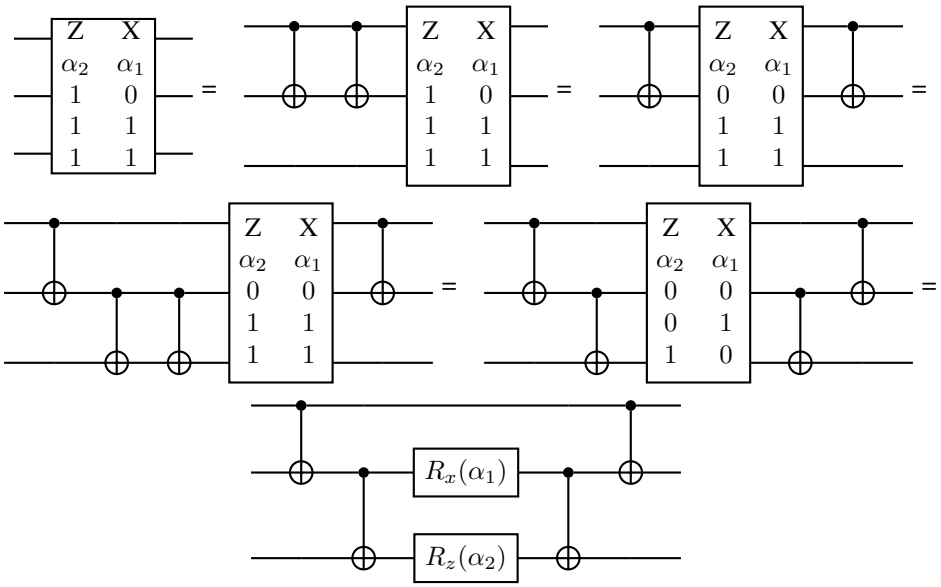

Fig. 1: Example of how the mixed ZX-phase polynomial intermediate representation can be used to generate a quantum circuit.

section. The simulated annealing algorithm [21] is used to carefully add a few CNOTs to the circuit and simplify the intermediate representation. However, for the final synthesis of the intermediate representation, they use a naive synthesis strategy.

Thus, for this paper, we replace the naive synthesis strategy with a smarter one, namely the algorithm from [24]. We will call this algorithm "paritysynth" since it tries to synthesize the circuit in an optimal parity ordering for the given connectivity graph. If a subsequence of phase gadgets mutually commutes (because they are of the same type), we can synthesize them in arbitrary order. Thus, we should find the order that is optimal for the CNOT count. This is what paritysynth does.

Lastly, once the intermediate representation in the circuit is empty, we are left with a long sequence of accumulated CNOTs after the black box. These CNOTs can then be optimized using the third algorithm: PermRowCol [17].

We will give a more in depth explanation of these algorithms in a future version of this paper. For now, we refer the interested reader to the original papers.

Add implementation details of these algorithms

explain Steiner trees [14]

The novelty of our work is that these algorithms have not yet been used in unison before nor compared against common transpilers like Qiskit and TKET. In the next section, we will see that this combination performs much better than the algorithms on their own.

IV. Results

Update the text to fit the new results in the tables.

To evaluate our transpilation procedure, we generated 100 random circuits containing m phase gadgets each acting on at least $\sqrt{n}$ (rounded) qubits, where n is the number of qubits.

The code generating the circuits and implementing our transpilation strategy can be found on Github [1]. Our results are generated using the Jupyter notebook "notebooks/6. Phase Circuit to Quantum Circuit.ipynb".

We compare our algorithm against our own implementation of the ParitySynth algorithm, the original implementation of the simulated annealing approach, the Qiskit transpiler, and the TKET compiler. For Qiskit, we use the normal "transpile()" method with optimization level 3. For TKET, we have recreated the default compilation pass with optimization level 2 from documentation [2]. For the simulated annealing we use 100 iterations and 5 CNOT blocks.

As target devices, we used common connectivity constraints of IBM quantum computers. The graphs representing these devices can be found in Figure 2. Note that the results for Valencia-like graphs are currently missing and will be added for the next version of this paper. We chose these devices to give an idea of the scalability of our approach and the effect of different real-world graphs on performance. We encourage readers who are interested in the performance for other graphs to adjust the provided Jupyter notebook for their use cases.

We show the average CNOT counts for our randomly generated circuits on the different connectivity graphs: 20-qubit devices in Figure **??**, a 14-qubit device in Figure **??**, and a 5-qubit device in Figure **??**. We show the average CNOT count for the originally generated, naively decomposed circuit (original), Qiskit, TKET, decomposed using only paritysynth (parity), naively decomposed and optimized by the annealer (annealer), and our combined approach (annealer+parity). Note that the results for our approach are not yet reallocating the qubits. A later version of this paper will have that, and the

[1] https://github.com/Aerylia/pauliopt

[2] https://cqcl.github.io/pytket-qiskit/api/index.html#default-compilation

	Original	Qiskit	TKET	ParitySynth	Annealer	Parity+RT	Parity+RT→Annealer
1	6.22	5.94	4.37	6.5	6.22	6.1	5.17
2	10.7	9.71	7.04	12.04	10.2	11.43	10.01
5	27.7	24.18	15.79	26.37	24.62	24.84	20.65
10	52.26	44.23	26.37	39.45	46.28	36.34	30.9
50	245.94	202.93	123.83	107.31	216.7	102.76	91.79
100	479.9	398.01	241.02	184.6	418.12	181.04	165.8

(a) 5-qubit Valencia-like devices

	Original	Qiskit	TKET	ParitySynth	Annealer	Parity+RT	Parity+RT→Annealer
1	4.82	4.58	3.9	4.82	4.9	4.65	4.7
2	9.28	8.51	6.71	8.87	9.14	8.63	8.17
5	23.46	20.2	14.23	17.89	19.32	17.62	15.57
10	43.98	37.79	25.36	29.53	37.84	27.1	24.42
50	205.54	171.92	114.9	77.37	155.98	72.91	67.71
100	413.5	346.77	225.63	137.68	304.16	132.67	123.12

(b) 5-qubit Yorktown-like devices

	Original	Qiskit	TKET	ParitySynth	Annealer	Parity+RT	Parity+RT→Annealer
1	17.64	17.34	13.34	18.48	18.34	18.26	18.17
2	35.88	34.77	26.38	36.99	35.9	36.7	35.2
5	89.48	82.74	62.08	88.22	87.86	88.12	80.9
10	184.3	171.89	124.57	179.18	177.68	178.86	164.76
50	914.1	846.19	609.22	605.71	848.58	582.8	555.66
100	1838.9	1698.72	1237.63	1029.33	1697.66	1007.96	968.02

(c) 14-qubit Melbourne-like devices

	Original	Qiskit	TKET	ParitySynth	Annealer	Parity+RT	Parity+RT→Annealer
1	26.66	26.41	19.59	28.3	27.42	28.07	27.44
2	53.56	52.25	38.26	54.56	54.0	54.37	52.58
5	137.46	133.38	92.19	142.22	133.28	142.22	133.0
10	276.44	266.32	185.27	277.88	268.26	277.88	259.34
50	1358.2	1300.9	926.38	1092.19	1302.44	1071.89	1028.61
100	2734.62	2621.78	1906.41	1785.13	2615.64	1757.1	1710.64

(d) 20-qubit Johannesburg-like devices

	Original	Qiskit	TKET	ParitySynth	Annealer	Parity+RT	Parity+RT→Annealer
1	29.0	28.71	22.08	30.88	29.6	30.38	29.6
2	56.58	54.95	39.34	60.77	56.0	60.64	56.69
5	138.5	132.72	91.68	149.28	134.2	149.28	138.18
10	278.28	263.14	182.07	282.2	267.72	282.2	258.54
50	1369.5	1281.56	931.5	1084.52	1297.54	1060.75	1011.73
100	2726.74	2562.69	1890.25	1779.63	2607.82	1754.11	1692.08

(e) 20-qubit Singapore-like devices

TABLE I: Average CNOT counts over 100 random circuits with different numbers of random phase gadgets and compiled for different devices. Original, Qiskit, and TKET represent the CNOT count in the original circuit and the count once compiled using Qiskit and TKET, respectively. ParitySynth and Annealer represent the two algorithms on which our work is based and added as a baseline. The last two columns are two proposed methods. Parity+RT is the ParitySynth algorithm with PermRowCol and Reverse Traversal. Parity+RT→Annealer also runs the annealer after the Reverse Traversal step.

reallocation will reduce the CNOT count.

It is clear that the addition of the annealer improves the paritysynth. For larger amounts of phase gadgets (≥ 50), the addition of paritysynth to the annealer also improves it.

We note that TKET is outperforming Qiskit, so we will focus our analysis for existing compilers to TKET's performance. TKET closely follows the performance of the annealer. If the circuit has many phase gadgets ($\geq$ 50, which is expected), our method does improve the CNOT count quite significantly with up to 50% for circuits with 100 phase gadgets.

V. Discussion

The proposed approach to the transpilation of quantum circuits relies on a few assumptions. In this section, we list these and discuss why we do not expect them to be a problem.

A. Original circuit quality

Our transpilation approach assumes that the original creator of the quantum circuit to be transpiled is not familiar with quantum circuit optimization and the qubit routing problem. Therefore, we assume that the given circuit is poorly optimized. This is a sensible assumption because current use cases for quantum computers is the simulation of quantum processes. Thus, the user of quantum computers are physicists and chemists, who might not be familiar with quantum computer science.

As such, we expect that most NISQ programs will be generated from a given Hamiltonian. During this process, the Hamiltonian is transformed into a sequence of Paulistrings, which make the mixed ZX-phase polynomial a native intermediate representation to the original program. In fact, using

	Qiskit	TKET	ParitySynth	Annealer	Parity+RT	Parity+RT→Annealer
1	0.0136	0.0941	0.0019	0.0065	0.0166	0.0548
2	0.0202	0.1507	0.003	0.008	0.0259	0.0776
5	0.0488	0.3648	0.0057	0.0122	0.0536	0.1505
10	0.0878	0.6018	0.0085	0.0148	0.083	0.2154
50	0.4479	2.9537	0.0334	0.0395	0.358	0.8247
100	0.8205	5.1633	0.067	0.0598	0.686	1.5587

(a) 5-qubit Valencia-like devices

	Qiskit	TKET	ParitySynth	Annealer	Parity+RT	Parity+RT→Annealer
1	0.0101	0.0816	0.0016	0.0067	0.015	0.0472
2	0.0175	0.1445	0.0027	0.0081	0.0247	0.071
5	0.0424	0.3243	0.0052	0.0125	0.0495	0.1341
10	0.085	0.6915	0.0094	0.0185	0.0918	0.2384
50	0.3503	2.6283	0.0293	0.0376	0.296	0.681
100	0.6688	4.8828	0.0611	0.0627	0.6099	1.3882

(b) 5-qubit Yorktown-like devices

	Qiskit	TKET	ParitySynth	Annealer	Parity+RT	Parity+RT→Annealer
1	0.0349	0.2801	0.0254	0.0092	0.2463	0.5923
2	0.0657	0.5536	0.0426	0.0129	0.4116	0.9826
5	0.1754	1.3947	0.0922	0.0255	0.9422	2.2392
10	0.3604	2.9009	0.1728	0.0438	1.7384	4.1142
50	2.2315	14.8995	0.7473	0.1667	7.3599	16.8116
100	4.9723	29.2199	1.5941	0.3326	16.0436	35.7677

(c) 14-qubit Melbourne-like devices

	Qiskit	TKET	ParitySynth	Annealer	Parity+RT	Parity+RT→Annealer
1	0.0511	0.4227	0.078	0.0128	0.7525	1.7707
2	0.0933	0.8298	0.1303	0.0199	1.2723	2.9239
5	0.2543	2.1881	0.309	0.046	3.0665	7.128
10	0.5594	4.6781	0.557	0.0839	5.5421	12.8963
50	3.2184	24.0644	2.6053	0.3639	25.8052	57.9768
100	7.139	49.2208	5.5031	0.7037	55.4527	121.9761

(d) 20-qubit Johannesburg-like devices

	Qiskit	TKET	ParitySynth	Annealer	Parity+RT	Parity+RT→Annealer
1	0.0547	0.4858	0.0857	0.0131	0.8504	1.9321
2	0.1018	0.8693	0.1339	0.0201	1.3207	3.0246
5	0.2565	2.1138	0.2927	0.0444	2.951	6.9142
10	0.5662	4.5812	0.5292	0.084	5.2872	12.343
50	3.5822	23.9334	2.4642	0.3604	24.6068	55.1975
100	7.831	46.1684	5.1464	0.6649	51.809	114.9429

(e) 20-qubit Singapore-like devices

TABLE II: Average runtime in seconds over 100 random circuits with different numbers of random phase gadgets and compiled for different devices. Original, Qiskit, and TKET represent the CNOT count in the original circuit and the count once compiled using Qiskit and TKET, respectively. ParitySynth and Annealer represent the two algorithms on which our work is based and added as a baseline. The last two columns are two proposed methods. Parity+RT is the ParitySynth algorithm with PermRowCol and Reverse Traversal. Parity+RT→Annealer also runs the annealer after the Reverse Traversal step.

Paulistrings as an intermediate representation was already proposed for the Paulihedral compiler [13].

We will mention that if the original quantum circuit is already heavily optimized for the target hardware, we do not expect our transpilation method to still improve the circuit. As we already could see in the results for simple circuits with few phase gadgets. The remainder of the paper will discuss why this could be the case and how one could deal with that.

B. The native multi-qubit interactions of the quantum computer

The only multi-qubit interaction that is generated by our algorithm is the CNOT gate. This makes sense from the point of view that together with arbitrary single qubit gates, this makes a universal gate set. However, the physical implementation of native multiqubit interactions on current quantum hardware is generally not a CNOT. The devices can execute a CNOT using their native interaction and some single qubit gates, but it is not immediately a CNOT. Some compilation is still necessary!

In general, it takes one of these multiqubit interactions to simulate the CNOT, making it seem that this does not affect the number of multiqubit interactions. However, some devices (e.g. ion-traps) allow multi-qubit interactions between all qubits. This would remove the need to reduce the phase gadgets into smaller interactions.

But let's assume, for the sake of argument, that we have a device with a native 2-qubit interaction. In that case, we can still simulate the CNOT using some single-qubit gates. However, on some devices these interactions will be implemented in an "echoed" form to improve environmental noise [6], doubling all 2-qubit gates. Moreover, in cases where a 2-qubit phase gadget acts on two connected qubits on the

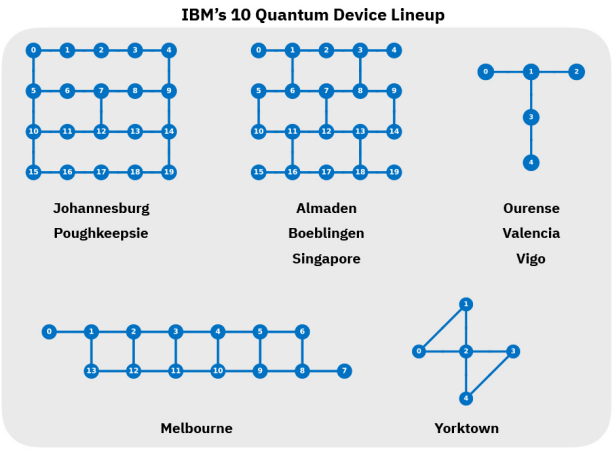

Fig. 2: Graphs representing the connectivity constraints for different IBM quantum computers. Picture taken from https://www.ibm.com/blogs/research/2019/09/quantum-computation-center/.

topology, the phase gadget might be turned into a native 2-qubit interaction directly (using single-qubit gates) without the need for generating CNOTs.

Transpiling a phase gadget directly to the native 2-qubit gate would require that 2-qubit gate to be tunable with the angle of the phase gadget. Some devices allow this (e.g. some IBM devices [6]), in which case they need to be calibrated for all possible angles rather than just the one angle needed for the CNOT (or two if it is echoed). But calibration of many angles is more difficult. So herein lies a trade-off: implement a 2-qubit phase gadget with 2 "good" interactions (from CNOTs) or 1 "bad" interaction (directly native)?

Our transpilation procedure will change the balance of this trade-off. Due to the resynthesis procedure, the second interaction of the decomposed phase gadget is pushed to the end of the circuit and might be optimized away. On the other hand, if the second interaction isn't optimized away, we might want to use the native interaction. But when the second interaction isn't pushed through the later phase gadget, it will change the entire circuit to the point that it is not obvious which option is better. What is clear in this trade-off is that, because the second interaction might be optimized away, the quality of the tunable 2-qubit interaction needs to be better than twice the fixed-angle interaction.

Additionally, our transpilation approach is not suitable for a look-up-and-replace strategy to introduce the tunable 2-qubit interactions, because it probably will not generate CNOT-(single qubit gate)-CNOT sequences. Thus, in case tunable interactions are allowed, the synthesis approach will need to take that into account from the start.

C. The native single qubit gates

The only single-qubit gates that our transpilation approach generates are X and Z rotations. The assumption here is that these are equally desirable for the hardware, but this is not necessarily true. For example, in superconducting devices, rotations around the Z axis are "virtual". They do not change the qubit itself but the equipment around it (i.e. it resets the clock). This means that Z rotations are essentially noiseless while X rotations require noise manipulations of the qubit. Thus, it is beneficial to generate more Z rotations that X rotations.

Given that X rotations will have to be calibrated and Z rotations do not, it is better to calibrate the X rotations for one or two angles and simulate all arbitrary X rotations $R_X(\alpha)$ as $HR_Z(\alpha)H$, where the H gate is created from the calibrated X rotations and the virtual Z rotation.

This adjustment can be immediately added to our proposed algorithm, but we can also make a slight adjustment to our algorithm that might influence the design of future quantum hardware. During the synthesis procedure, we alternate between decomposing sequences of Z-phase gadgets and X-phase gadgets. What we can do is to generate a H gate on every qubit before the sequence of X-phase gadgets and after the sequence of X-phase gadgets. Due to the connectivity constraints, we do no know which qubits are needed to synthesize the X-phase gadgets, so we need to put the H gate on all qubits (unnecessary H gates can be removed after synthesis). Now, the sequence of X-phase gadgets is turned into a sequence of Z-phase gadgets, and we can continue the algorithm like before, except that when a generated CNOT is pushed through the barrier of H gates, its direction is reversed. In case that the sequence of X-phase gadgets is longer than the number of qubits, this reduces the total amount of H gates required.

The reason we believe strategy might influence the design of future quantum hardware is that placing a H gate on all qubits is a very common operation in quantum algorithms [7, 22, 10]. If the transpilation also generates these frequently, it might be sensible to make this operation "hardware accelerated". I.e. make some hardware adjustments so that this specific thing can be done better/faster/easier.

D. Effect of gate fidelity

Another effect that our algorithm does not natively take into account is the difference in gate fidelity. On current quantum computers, not every qubit can be equally well controlled and not every multi-qubit interaction has the same noise. Our results do not take this into account and assume that all gates are equally good. We have discussed some of these aspects when discussing the implementation of the natively supported gates by the quantum devices, but we will discuss the effect of fidelity differences in a broader sense here.

Although we assume for our results that all gates are equally good, our algorithm can be adjusted to take gate quality into account. The most straightforward way to do this is to make the connectivity graph a weighted graph where the weights correspond to the gate fidelities. More generally, we can redefine the measure of "qubit distance" when calculating the minimal Steiner trees. Within our implementation, this would

mean to redefine the value of the shortest path between every qubit in the topology. Since we approximate the Steiner-tree by calculating the minimal spanning tree over the all-pairs shortest paths. This measure of qubit distance can then take more into account than only the 2-qubit gate fidelity, but also the quality of the qubit itself. Similarly, we can use gate fidelity as a tie-breaker when choosing CNOTs or qubits in our algorithm.

In general, we need to find better strategies to take more dynamic sources of noise into account. The algorithm can be adjusted to reduce crosstalk between qubits, improve gate scheduling, or take into account other device-specific sources of error. This will require more collaboration between hardware manufacturers and transpiler designers, as well as availability of more detailed specifications for the available hardware.

E. Effect of approximate Steiner-trees

Lastly, we discuss the elephant in the room: finding a minimal Steiner tree is an NP-hard problem [14]. The only reason our algorithm is performant is that we use a polynomial approximation. This means that the paths that we find between qubits might not be optimal and we could reduce the CNOT count even further.

It is possible that the types of connectivity graphs on quantum computers are sparse enough so that the approximation is able to find the minimum Steiner tree, but we do not know.

In case the approximate Steiner trees are insufficient, we could still find a minimal Steiner tree using quantum computers. Thanks to adiabetic quantum computing, we should be able to find minimal Steiner trees with quantum annealers [15]. In theory, we could also use gate-based quantum computers to do this, but we run into a chicken-egg problem. Because current quantum annealers have more qubits to their disposal and they do not need a transpiled quantum circuit, we can actually use quantum computers to compile quantum circuits. In fact, this has recently already been done in the ISAAQ compiler [19].

More generally, we might be able to describe the full parity network problem (solved in paritysynth) in a fashion suitable for quantum annealers. There is a lot of interaction between generating which parity on which qubit in which order, that might be very suitable for a quantum annealing formulation.

VI. Conclusion and Future work

In this paper, we proposed a full-stack quantum circuit resynthesis approach to transpilation. We compared it to similar state-of-the-art algorithms as well as the Qiskit transpiler and TKET. We showed that for more complicated circuits our proposed approach improves the CNOT count.

We argue that in transpilation, we need to take a more holistic approach that takes the semantics of the circuit into account, not just the gates themselves. And we believe that the improved CNOT counts that we have shown support that argument.

However, our algorithm can still be improved. The fact that the simulated annealing is improving the synthesized circuit shows that our heuristics are non-optimal. For now, this is good enough, but we should be able to do better.

Additionally, we need to generalize our algorithm to have Paulistrings as our intermediate representation, just like the Paulihedral compiler [13]. Transforming the Paulistrings into a mixed ZX-phase polynomial can generate many extra phase gadgets which is inefficient.

Lastly, one of the strengths of compiling methods like Qiskit and TKET is that they are able to move the qubits on the architecture. The implementation for the current results does not choose an optimal mapping of the qubit, nor does it move the qubits around. This means that improvements can still be made, as shown in [17]. The next version of this paper will include the dynamic reallocation from PermRowCol [17] and hopefully the reverse traversal strategy from [12] to find an optimal mapping if time allows. Regardless, applying these methods will only further improve the results.

Acknowledgement

I would like to thank Richie Yeung and Stefano Gogioso for helpful discussions and insights into their code base on which the full algorithm was built.

References

[1] A-tA-v et al. *Qiskit: An Open-source Framework for Quantum Computing*. 2021. DOI: 10.5281/zenodo.2573505.

[2] Amazon Web Services. *Amazon Braket*. 2023. URL: https://aws.amazon.com/braket/.

[3] Matthew Amy, Parsiad Azimzadeh, and Michele Mosca. "On the controlled-NOT complexity of controlled-NOT–phase circuits". en. In: *Quantum Science and Technology* 4.1 (Sept. 2018), p. 015002. ISSN: 2058-9565. DOI: 10.1088/2058-9565/aad8ca.

[4] Matthew Amy, Dmitri Maslov, and Michele Mosca. "Polynomial-Time T-Depth Optimization of Clifford+T Circuits Via Matroid Partitioning". In: *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 33.10 (Oct. 2014), pp. 1476–1489. ISSN: 1937-4151. DOI: 10.1109/TCAD.2014.2341953.

[5] Alexander Cowtan et al. "On the Qubit Routing Problem". en. In: (2019), 32 pages. DOI: 10.4230/LIPICS.TQC.2019.5.

[6] Nathan Earnest, Caroline Tornow, and Daniel J. Egger. "Pulse-Efficient Circuit Transpilation for Quantum Applications on Cross-Resonance-Based Hardware". In: *Physical Review Research* 3.4 (Oct. 2021), p. 043088. ISSN: 2643-1564. DOI: 10.1103/PhysRevResearch.3.043088. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.3.043088 (visited on 10/03/2022).

[7] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. "A Quantum Approximate Optimization Algorithm". In: (2014). DOI: 10.48550/ARXIV.1411.4028. URL: https://arxiv.org/abs/1411.4028.

[8] Vlad Gheorghiu et al. “Reducing the CNOT Count for Clifford+T Circuits on NISQ Architectures”. In: (2020). DOI: 10.48550/ARXIV.2011.12191. URL: https://arxiv.org/abs/2011.12191 (visited on 10/03/2022).

[9] Stefano Gogioso and Richie Yeung. “Annealing Optimisation of Mixed ZX Phase Circuits”. In: (2022). DOI: 10.48550/ARXIV.2206.11839. URL: https://arxiv.org/abs/2206.11839 (visited on 10/03/2022).

[10] Lov K. Grover. “A fast quantum mechanical algorithm for database search”. en. In: *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing - STOC ’96*. Philadelphia, Pennsylvania, United States: ACM Press, 1996, pp. 212–219. ISBN: 978-0-89791-785-8. DOI: 10.1145/237814.237866. URL: http://portal.acm.org/citation.cfm?doid=237814.237866.

[11] Aleks Kissinger and Arianne Meijer van de Griend. “CNOT Circuit Extraction for Topologically-Constrained Quantum Memories”. In: *Quantum Information and Computation* 20.7&8 (June 2020), pp. 581–596. ISSN: 15337146, 15337146. DOI: 10.26421/QIC20.7-8-4. URL: http://www.rintonpress.com/journals/doi/QIC20.7-8-4.html (visited on 10/03/2022).

[12] Gushu Li, Yufei Ding, and Yuan Xie. “Tackling the qubit mapping problem for NISQ-era quantum devices”. In: *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*. 2019, pp. 1001–1014.

[13] Gushu Li et al. “Paulihedral: A Generalized Block-Wise Compiler Optimization Framework for Quantum Simulation Kernels”. In: *Proceedings of the 27th ACM International Conference on Architectural Support for Programming Languages and Operating Systems*. Lausanne Switzerland: ACM, Feb. 2022, pp. 554–569. ISBN: 978-1-4503-9205-1. DOI: 10.1145/3503222.3507715. URL: https://dl.acm.org/doi/10.1145/3503222.3507715 (visited on 10/03/2022).

[14] Ivana Ljubić. “Solving Steiner trees: Recent advances, challenges, and perspectives”. en. In: *Networks* 77.2 (Mar. 2021), pp. 177–204. ISSN: 0028-3045, 1097-0037. DOI: 10.1002/net.22005.

[15] Andrew Lucas. “Ising formulations of many NP problems”. In: *Frontiers in Physics* 2 (2014). ISSN: 2296-424X. DOI: 10.3389/fphy.2014.00005. URL: http://journal.frontiersin.org/article/10.3389/fphy.2014.00005/abstract.

[16] Adriana Meijer - van de Griend and Ross Duncan. “Architecture-Aware Synthesis of Phase Polynomials for NISQ Devices”. In: (2020). DOI: 10.48550/ARXIV.2004.06052. URL: https://arxiv.org/abs/2004.06052 (visited on 10/03/2022).

[17] Adriana Meijer - van de Griend and Sarah Meng Li. “Dynamic Qubit Allocation and Routing for Constrained Topologies by CNOT Circuit Re-Synthesis”. In: (2022). DOI: 10.48550/ARXIV.2205.00724. URL: https://arxiv.org/abs/2205.00724 (visited on 10/03/2022).

[18] Prakash Murali et al. “Formal constraint-based compilation for noisy intermediate-scale quantum systems”. en. In: *Microprocessors and Microsystems* 66 (Apr. 2019), pp. 102–112. ISSN: 01419331. DOI: 10.1016/j.micpro.2019.02.005.

[19] Soshun Naito et al. “ISAAQ: Ising Machine Assisted Quantum Compiler”. In: *arXiv preprint arXiv:2303.02830* (2023).

[20] Beatrice Nash, Vlad Gheorghiu, and Michele Mosca. “Quantum circuit optimizations for NISQ architectures”. In: *Quantum Science and Technology* 5.2 (2020), p. 025010.

[21] Martin Pincus. “A Monte Carlo Method for the Approximate Solution of Certain Types of Constrained Optimization Problems”. en. In: *Operations Research* 18.6 (Dec. 1970), pp. 1225–1228. ISSN: 0030-364X, 1526-5463. DOI: 10.1287/opre.18.6.1225.

[22] P.W. Shor. “Algorithms for quantum computation: discrete logarithms and factoring”. In: *Proceedings 35th Annual Symposium on Foundations of Computer Science*. Santa Fe, NM, USA: IEEE Comput. Soc. Press, 1994, pp. 124–134. ISBN: 978-0-8186-6580-6. DOI: 10.1109/SFCS.1994.365700. URL: http://ieeexplore.ieee.org/document/365700/.

[23] Seyon Sivarajah et al. “t—ket⟩: a retargetable compiler for NISQ devices”. In: 6 (Jan. 2021), p. 014003. ISSN: 2058-9565. DOI: 10.1088/2058-9565/ab8e92.

[24] Vivien Vandaele, Simon Martiel, and Timothée Goubault de Brugière. “Phase Polynomials Synthesis Algorithms for NISQ Architectures and Beyond”. In: *Quantum Science and Technology* 7.4 (Oct. 2022), p. 045027. ISSN: 2058-9565. DOI: 10.1088/2058-9565/ac5a0e. URL: https://iopscience.iop.org/article/10.1088/2058-9565/ac5a0e (visited on 09/29/2022).

[25] Davide Venturelli et al. “Compiling quantum circuits to realistic hardware architectures using temporal planners”. In: *Quantum Science and Technology* 3.2 (Apr. 2018), p. 025004. ISSN: 2058-9565. DOI: 10.1088/2058-9565/aaa331.

[26] John van de Wetering. “ZX-calculus for the working quantum computer scientist”. In: *arXiv preprint arXiv:2012.13966* (2020).

[27] William K Wootters and Wojciech H Zurek. “A single quantum cannot be cloned”. In: *Nature* 299 (1982), pp. 802–803.